

CLAIMS

We claim:

1. A system for matching one or more abbreviations and one or more definitions, comprising:

an abbreviation pattern generation process that generates one or more abbreviation patterns corresponding to the candidate abbreviations; and

a definition pattern generation process that generates one or more definition patterns corresponding to the candidate definitions.

2. A system, as in claim 1, further comprising:

a set of abbreviation rules that correlate abbreviation patterns to definition patterns using one or more formation rules;

a lookup process that selects one or more formation rules, being selected formation rules, corresponding to the abbreviation pattern of the candidate abbreviation and the definition pattern of the candidate definition; and

a rule application process that applies the selected formation rules to determine which candidate definitions match the candidate abbreviation.

3. A system, as in claim 1, further comprising:

one or more matching algorithms that match one or more pairs of abbreviations and definitions based on the abbreviation patterns and the definition patterns.

4. A system, as in claim 2, further comprising:

one or more matching algorithms that match one or more pairs of abbreviations and definitions based on the abbreviation patterns and the definition patterns.

5. A system, as in claim 4, where rule application process and the matching algorithm apply both rule based and non rule based matching processes to match one or more abbreviations and one or more definitions.

10 6. A system, as in claim 1, further comprising:

a method for specifying pairs, each of which contains a candidate abbreviation and a candidate definition, for each pair generating an abbreviation patterns and a definition pattern.

7. A system, as in claim 6, where the pairs include one or more of the following:

- a user input pair,

- an existing abbreviation database pair, and
- a pair produced by a program interface.

8. A system, as in claim 1, further comprising:

an abbreviation recognition process that finds one or more candidate abbreviations in text.

9. A system, as in claim 1, further comprising:

a definition finding process that locates one or more candidate definitions corresponding to the candidate abbreviation.

10. A system, as in claim 1, further comprising:

a best match selection process that chooses a best candidate definition from the matched candidate definitions using one or more criteria.

11. A system, as in claim 10, further comprising:

a best match selection mechanism that employs one or more weighting features.

12. A system, as in claim 11, where the weighting features may include one or more of the followings:

- syntactic cues found in the context ,
- rule priority of the formation rule that matched the pair,
- the distance of the abbreviaiton and the defintion,
- capitalization of the definition,
- number of words in the definition, and
- number of stopwords in the definition.

5

13. A system, as in claim 1, further comprising:

an output process the outputs the candidate abbreviation and the matched candidate definition as confirmed pairs.

14. A system, as in claim 2, where the formation rule that produced the best candidate definition is weighted better due to the choice of the best candidate definition.

15. A system, as in claim 2, further comprising:

a process for adding new abbreviation rules in the existing set of abbreviation rules.

16. A system, as in claim 2, further comprising:

a mechanism for generating one or more new abbreviation rules when no formation rules successfully match high-quality pairs of candidate abbreviations and definitions.

17. A system, as in claim 16, further comprising:

a process for automatically adding the generated abbreviation rules to the existing set of
5 abbreviation rules.

18. A system, as in claim 2, further comprising:

a rule generation process for generating abbreviation rules from pairs of abbreviations and
definitions.

19. A system, as in claim 1, further comprising:

a set of layered matching algorithms which are based on the relationship between the lengths of
abbreviation patterns and the lengths of definition patterns.

20. A system, as in claim 19, where each algorithm in the layered matching mechanism is applied
in priority sequence.

21. A system for matching one or more abbreviations and one or more definitions, comprising:

means for generating one or more abbreviation patterns corresponding to the candidate abbreviations; and

means for generating one or more definition patterns corresponding to the candidate definitions.

22. A method for matching one or more abbreviations and one or more definitions, comprising:

generating one or more abbreviation patterns corresponding to the candidate abbreviations; and

generating one or more definition patterns corresponding to the candidate definitions.